Minimizing Weighted Tree Grammars using Simulation

Andreas Maletti

Pretoria — July 24, 2009

Author supported by MEC grants JDCI-2007-760 and MTM-2007-63422.

Overview

Simulations

- "Half" a bisimulation
- Used in (lossless) reductions of the size of grammars
- Also used in logic (inseparability by logic formulae)

Rough definition

A nonterminal B simulates another nonterminal A if any rule involving A is covered by a rule involving B

But

our new definition is not as intuitive

Overview

Simulations

- “Half” a bisimulation
- Used in (lossless) reductions of the size of grammars
- Also used in logic (inseparability by logic formulae)

Rough definition

A nonterminal B simulates another nonterminal A if any rule involving A is covered by a rule involving B

But

our new definition is not as intuitive

Overview

Simulations

- "Half" a bisimulation
- Used in (lossless) reductions of the size of grammars
- Also used in logic (inseparability by logic formulae)

Rough definition

A nonterminal B simulates another nonterminal A if any rule involving A is covered by a rule involving B

But

our new definition is not as intuitive

State of the art

Property	*Existing Simulation* Backward	Forward
Generalization		
unweighted simulation	✓	✓
weighted bisimulation	✗	✗
Computation		
Admits greatest simulation	✓	✓
Yields minimal device	✗	✗
Easy to compute	✓	✓

State of the art

Property	*Existing Simulation* Backward	Forward
Generalization		
unweighted simulation	✓	✓
weighted bisimulation	✗	✗
Computation		
Admits greatest simulation	✓	✓
Yields minimal device	✗	✗
Easy to compute	✓	✓

Table of contents

Overview

Tree Series

- Assigns weight (e.g. a probability) to each tree
- Weights drawn from semiring; e.g. $([0,1], \max, \cdot, 0, 1)$

Weighted Tree Grammar

- Finite representation of a tree series

Applications

- Re-ranker implementing a language model
- Representation of a parser (or parses)

Overview

Tree Series

- Assigns weight (e.g. a probability) to each tree
- Weights drawn from semiring; e.g. $([0, 1], \max, \cdot, 0, 1)$

Weighted Tree Grammar

- Finite representation of a tree series

Applications

- Re-ranker implementing a language model
- Representation of a parser (or parses)

Overview

Tree Series

- Assigns weight (e.g. a probability) to each tree
- Weights drawn from semiring; e.g. $([0, 1], \max, \cdot, 0, 1)$

Weighted Tree Grammar

- Finite representation of a tree series

Applications

- Re-ranker implementing a language model
- Representation of a parser (or parses)

Definition

Weighted tree grammar (WTG) is tuple $(N, \Sigma, \mathcal{S}, S, P)$ where

- N: finite set of *nonterminals*
- Σ: ranked alphabet of *symbols*
- $\mathcal{S} = (\mathcal{S}, +, \cdot, 0, 1)$: semiring of *weights*
- $S \in N$: *start nonterminal*
- P finite set of productions of the form $A \xrightarrow{s} \sigma(A_1, \ldots, A_k)$ with $A, A_1, \ldots, A_k \in N$, $s \in \mathcal{S}$, and $\sigma \in \Sigma_k$

Notation

We write $\mathrm{wt}(A \to \sigma(A_1, \ldots, A_k)) = s$ if $A \xrightarrow{s} \sigma(A_1, \ldots, A_k) \in P$

Syntax

Definition

Weighted tree grammar (WTG) is tuple $(N, \Sigma, \mathcal{S}, S, P)$ where

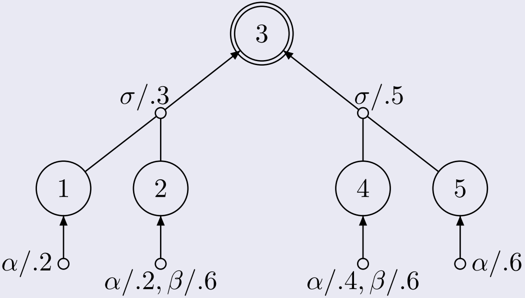
- N: finite set of *nonterminals*
- Σ: ranked alphabet of *symbols*
- $\mathcal{S} = (\mathcal{S}, +, \cdot, 0, 1)$: semiring of *weights*
- $S \in N$: *start nonterminal*
- P finite set of productions of the form $A \xrightarrow{s} \sigma(A_1, \dots, A_k)$ with $A, A_1, \dots, A_k \in N$, $s \in \mathcal{S}$, and $\sigma \in \Sigma_k$

Notation

We write $\mathrm{wt}(A \to \sigma(A_1, \dots, A_k)) = s$ if $A \xrightarrow{s} \sigma(A_1, \dots, A_k) \in P$

Syntax — Illustration

Sample Grammar



Semantics

Restriction

Here only for idempotent (i.e. $1 + 1 = 1$) semirings

Definition

Let $c \in C_\Sigma(N)$ and $p = A \xrightarrow{s} \sigma(A_1, \dots, A_k) \in P$

$$c[A] \stackrel{p}{\Rightarrow} c[\sigma(A_1, \dots, A_k)]$$

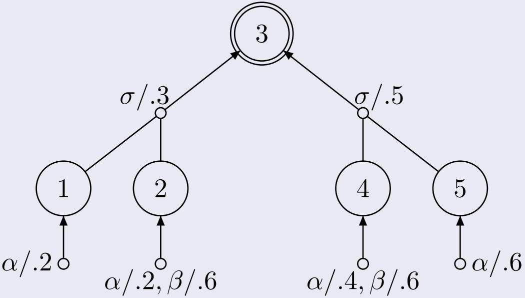
Weight of a tree $t \in T_\Sigma$ in nonterminal A:

$$\mathrm{wt}(t, A) = \sum_{\substack{p_1, \dots, p_n \in P \\ A \stackrel{p_1}{\Rightarrow} \cdots \stackrel{p_n}{\Rightarrow} t}} \mathrm{wt}(p_1) \cdot \ldots \cdot \mathrm{wt}(p_n)$$

Weight of t: $\mathrm{wt}(t) = \mathrm{wt}(t, S)$

Semantics — Illustration

Sample Grammar



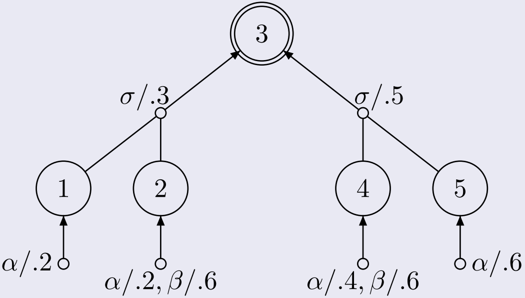
Sample derivations for input tree: $\sigma(\alpha, \alpha)$

$3 \Rightarrow_G \sigma(1, 2) \Rightarrow_G \sigma(\alpha, 2) \Rightarrow_G \sigma(\alpha, \alpha)$ rule weights: $0.3 \cdot 0.2 \cdot 0.2$

$3 \Rightarrow_G \sigma(4, 5) \Rightarrow_G \sigma(\alpha, 5) \Rightarrow_G \sigma(\alpha, \alpha)$ rule weights: $0.5 \cdot 0.4 \cdot 0.6$

Semantics — Illustration

Sample Grammar



Sample derivations for input tree: $\sigma(\alpha, \alpha)$

$3 \Rightarrow_G \sigma(1, 2) \Rightarrow_G \sigma(\alpha, 2) \Rightarrow_G \sigma(\alpha, \alpha)$ rule weights: $0.3 \cdot 0.2 \cdot 0.2$

$3 \Rightarrow_G \sigma(4, 5) \Rightarrow_G \sigma(\alpha, 5) \Rightarrow_G \sigma(\alpha, \alpha)$ rule weights: $0.5 \cdot 0.4 \cdot 0.6$

Idempotent semiring

Definition

Idempotent semiring: $1 + 1 = 1$

Definition

Natural order: $a \sqsubseteq b$ if and only if $a + b = b$

Idempotent semiring

Definition

Idempotent semiring: $1 + 1 = 1$

Definition

Natural order: $a \sqsubseteq b$ if and only if $a + b = b$

Table of contents

Backward simulation

Old definition

A quasi-order $\preceq \subseteq N^2$ is a backward simulation if for every $A \preceq B$, $\sigma \in \Sigma_k$, and $A_1, \ldots, A_k \in N$ there exist $A_1 \preceq B_1, \ldots, A_k \preceq B_k$ such that

$$\mathrm{wt}(A \to \sigma(A_1, \ldots, A_k)) \sqsubseteq \mathrm{wt}(B \to \sigma(B_1, \ldots, B_k))$$

New definition

A quasi-order $\preceq \subseteq N^2$ is a backward simulation if for every $A \preceq B$, $\sigma \in \Sigma_k$, and $A_1, \ldots, A_k \in N$

$$\sum_{\substack{B_1 \in \uparrow A_1 \\ \cdots \\ B_k \in \uparrow A_k}} \mathrm{wt}(A \to \sigma(B_1, \ldots, B_k)) \sqsubseteq \sum_{\substack{B_1 \in \uparrow A_1 \\ \cdots \\ B_k \in \uparrow A_k}} \mathrm{wt}(B \to \sigma(B_1, \ldots, B_k))$$

Backward simulation

Old definition

A quasi-order $\preceq \subseteq N^2$ is a backward simulation if for every $A \preceq B$, $\sigma \in \Sigma_k$, and $A_1, \ldots, A_k \in N$ there exist $A_1 \preceq B_1, \ldots, A_k \preceq B_k$ such that

$$\mathrm{wt}(A \to \sigma(A_1, \ldots, A_k)) \sqsubseteq \mathrm{wt}(B \to \sigma(B_1, \ldots, B_k))$$

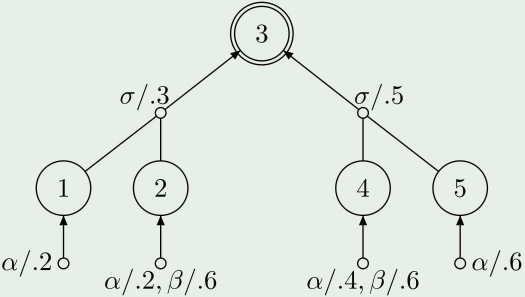
New definition

A quasi-order $\preceq \subseteq N^2$ is a backward simulation if for every $A \preceq B$, $\sigma \in \Sigma_k$, and $A_1, \ldots, A_k \in N$

$$\sum_{\substack{B_1 \in \uparrow A_1 \\ \vdots \\ B_k \in \uparrow A_k}} \mathrm{wt}(A \to \sigma(B_1, \ldots, B_k)) \sqsubseteq \sum_{\substack{B_1 \in \uparrow A_1 \\ \vdots \\ B_k \in \uparrow A_k}} \mathrm{wt}(B \to \sigma(B_1, \ldots, B_k))$$

An example

Example



$1 \preceq 2 \preceq 4$ and $1 \preceq 5$

Preliminary results

Theorem

There exists a greatest backward simulation for G.

Lemma (Main lemma)

For every $t \in T_\Sigma$ and $A \preceq B$

$$\mathrm{wt}(t, A) \sqsubseteq \mathrm{wt}(t, B)$$

Corollary

For every $t \in T_\Sigma$ and $A \preceq B \preceq A$

$$\mathrm{wt}(t, A) = \mathrm{wt}(t, B)$$

Preliminary results

Theorem

There exists a greatest backward simulation for *G*.

Lemma (Main lemma)

For every $t \in T_\Sigma$ and $A \preceq B$

$$\mathrm{wt}(t, A) \sqsubseteq \mathrm{wt}(t, B)$$

Corollary

For every $t \in T_\Sigma$ and $A \preceq B \preceq A$

$$\mathrm{wt}(t, A) = \mathrm{wt}(t, B)$$

Preliminary results

Theorem

There exists a greatest backward simulation for G.

Lemma (Main lemma)

For every $t \in T_\Sigma$ *and* $A \preceq B$

$$\mathrm{wt}(t, A) \sqsubseteq \mathrm{wt}(t, B)$$

Corollary

For every $t \in T_\Sigma$ *and* $A \preceq B \preceq A$

$$\mathrm{wt}(t, A) = \mathrm{wt}(t, B)$$

Reducing the grammar

Definition

Let $\simeq = \preceq \cap \preceq^{-1}$. Let $(G/\simeq) = (N', \Sigma, \mathcal{S}, S', P')$ with

- $N' = (N/\simeq)$
- $S' = [S]$
- for every $\sigma \in \Sigma_k$ and $A', A'_1, \ldots, A'_k \in N'$

$$\text{wt}(A' \to \sigma(A'_1, \ldots, A'_k)) = \sum_{A_1 \in A'_1, \ldots, A_k \in A'_k} \text{wt}(A \to \sigma(A_1, \ldots, A_k))$$

with $A \in A'$

Theorem

G and $(G/\simeq)$ are equivalent.

Reducing the grammar

Definition

Let $\simeq \, = \, \preceq \cap \preceq^{-1}$. Let $(G/\simeq) = (N', \Sigma, \mathcal{S}, S', P')$ with

- $N' = (N/\simeq)$
- $S' = [S]$
- for every $\sigma \in \Sigma_k$ and $A', A'_1, \ldots, A'_k \in N'$

$$\text{wt}(A' \to \sigma(A'_1, \ldots, A'_k)) = \sum_{A_1 \in A'_1, \ldots, A_k \in A'_k} \text{wt}(A \to \sigma(A_1, \ldots, A_k))$$

with $A \in A'$

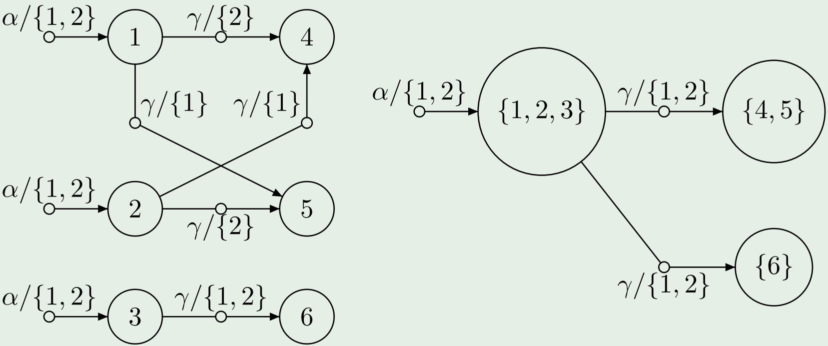
Theorem

G and $(G/\simeq)$ are equivalent.

Full example

Semiring $(\mathcal{P}(\{1,2\}), \cup, \cap, \emptyset, \{1,2\})$

Example (Old backward simulation)

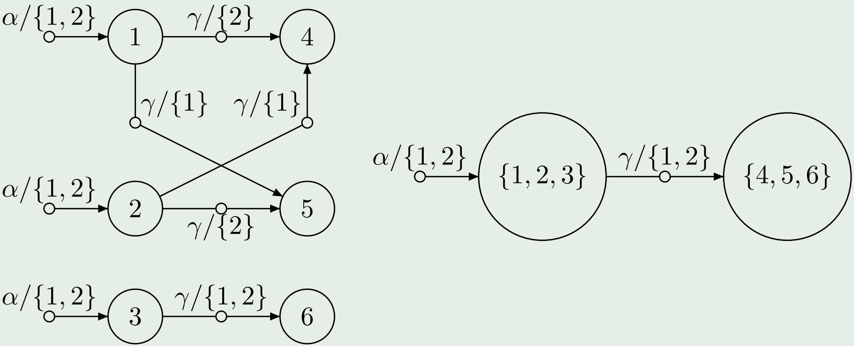


$1 \preceq 2 \preceq 3 \preceq 1$ and $4 \preceq 5 \preceq 4 \preceq 6$, but $6 \not\preceq 4$

Full example

Semiring $(\mathcal{P}(\{1,2\}), \cup, \cap, \emptyset, \{1,2\})$

Example (New backward simulation)



$1 \preceq 2 \preceq 3 \preceq 1$ and $4 \preceq 5 \preceq 6 \preceq 4$

Algorithm computing the greatest simulation

$R_0 \leftarrow N \times N$
$i \leftarrow 0$
repeat
  $j \leftarrow i$
  for all $\sigma \in \Sigma_k$ and $A_1, \ldots, A_k \in N$ **do**
    $R_{i+1} \leftarrow \{(A, B) \in R_i \mid$

$$\sum_{\substack{B_1 \in \uparrow_{R_i} A_1 \\ \vdots \\ B_k \in \uparrow_{R_i} A_k}} \mathrm{wt}(A \to \sigma(B_1, \ldots, B_k)) \sqsubseteq \sum_{\substack{B_1 \in \uparrow_{R_i} A_1 \\ \vdots \\ B_k \in \uparrow_{R_i} A_k}} \mathrm{wt}(B \to \sigma(B_1, \ldots, B_k))\}$$

    $i \leftarrow i + 1$
  end for
until $R_i = R_j$

Additional (new) properties

Theorem

Every backward bisimulation [Högberg et al '07] is a backward simulation.

Theorem

If $\preceq$ is the greatest backward simulation, then $(G/\simeq)$ with $\simeq = \preceq \cap \preceq^{-1}$ is backward-simulation minimal.

(i.e., it cannot be reduced further via backward simulation)

Additional (new) properties

Theorem

Every backward bisimulation [Högberg et al '07] is a backward simulation.

Theorem

If $\preceq$ *is the greatest backward simulation, then* $(G/\simeq)$ *with* $\simeq \, = \, \preceq \cap \preceq^{-1}$ *is backward-simulation minimal.*

(i.e., it cannot be reduced further via backward simulation)

State of the art

Property	*Simulation* New backward	Old forward
Generalization		
unweighted simulation	✓	✓
weighted bisimulation	✓	✗
Computation		
Admits greatest simulation	✓	✓
Yields minimal device	✓	✗
Easy to compute	?	✓

Table of contents

Forward simulation

Old definition

A quasi-order $\preceq \subseteq N^2$ is a forward simulation if for every $A \preceq B$

- if $A = S$, then $B = S$, and
- for every $\sigma \in \Sigma_k$, $1 \leq i \leq k$, and $A', A_1, \ldots, A_k \in N$ there exist $B' \in N$ such that $A' \preceq B'$ and

$$\mathrm{wt}(A' \to \sigma(A_1, \ldots, A, \ldots, A_k)) \sqsubseteq \mathrm{wt}(B' \to \sigma(A_1, \ldots, B, \ldots, A_k))$$

New definition

A quasi-order $\preceq \subseteq N^2$ is a forward simulation if for every $A \preceq B$

- if $A = S$, then $B = S$, and
- for every $\sigma \in \Sigma_k$, $1 \leq i \leq k$, and $A', A_1, \ldots, A_k \in N$

$$\sum_{B' \in \uparrow A'} \mathrm{wt}(B' \to \sigma(\ldots, A, \ldots)) \sqsubseteq \sum_{B' \in \uparrow A'} \mathrm{wt}(B' \to \sigma(\ldots, B, \ldots))$$

Forward simulation

Old definition

A quasi-order $\preceq \subseteq N^2$ is a forward simulation if for every $A \preceq B$

- if $A = S$, then $B = S$, and
- for every $\sigma \in \Sigma_k$, $1 \leq i \leq k$, and $A', A_1, \ldots, A_k \in N$ there exist $B' \in N$ such that $A' \preceq B'$ and

$$\mathrm{wt}(A' \to \sigma(A_1, \ldots, A, \ldots, A_k)) \sqsubseteq \mathrm{wt}(B' \to \sigma(A_1, \ldots, B, \ldots, A_k))$$

New definition

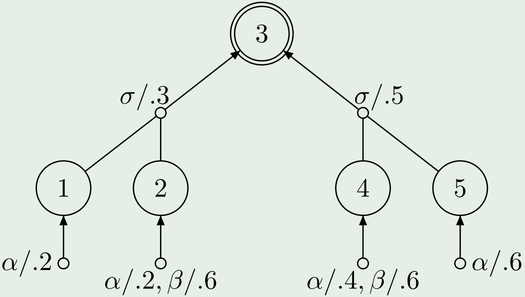
A quasi-order $\preceq \subseteq N^2$ is a forward simulation if for every $A \preceq B$

- if $A = S$, then $B = S$, and
- for every $\sigma \in \Sigma_k$, $1 \leq i \leq k$, and $A', A_1, \ldots, A_k \in N$

$$\sum_{B' \in \uparrow A'} \mathrm{wt}(B' \to \sigma(\ldots, A, \ldots)) \sqsubseteq \sum_{B' \in \uparrow A'} \mathrm{wt}(B' \to \sigma(\ldots, B, \ldots))$$

An example

Example



no similar states

Preliminary results

Theorem

There exists a greatest forward simulation for G.

Lemma (Main lemma)

For every $c \in C_\Sigma$, $A \preceq B$, and up-set $U \subseteq N$

$$\sum_{B' \in U} \mathrm{wt}(c[A], B') \sqsubseteq \sum_{B' \in U} \mathrm{wt}(c[B], B')$$

Corollary

For every $c \in C_\Sigma$, $A \preceq B \preceq A$, and up-set $U \subseteq N$

$$\sum_{B' \in U} \mathrm{wt}(c[A], B') = \sum_{B' \in U} \mathrm{wt}(c[B], B')$$

Preliminary results

Theorem

There exists a greatest forward simulation for G.

Lemma (Main lemma)

For every $c \in C_\Sigma$, $A \preceq B$, and up-set $U \subseteq N$

$$\sum_{B' \in U} \mathrm{wt}(c[A], B') \sqsubseteq \sum_{B' \in U} \mathrm{wt}(c[B], B')$$

Corollary

For every $c \in C_\Sigma$, $A \preceq B \preceq A$, and up-set $U \subseteq N$

$$\sum_{B' \in U} \mathrm{wt}(c[A], B') = \sum_{B' \in U} \mathrm{wt}(c[B], B')$$

Preliminary results

Theorem

There exists a greatest forward simulation for G.

Lemma (Main lemma)

For every $c \in C_\Sigma$, $A \preceq B$, and up-set $U \subseteq N$

$$\sum_{B' \in U} \mathrm{wt}(c[A], B') \sqsubseteq \sum_{B' \in U} \mathrm{wt}(c[B], B')$$

Corollary

For every $c \in C_\Sigma$, $A \preceq B \preceq A$, and up-set $U \subseteq N$

$$\sum_{B' \in U} \mathrm{wt}(c[A], B') = \sum_{B' \in U} \mathrm{wt}(c[B], B')$$

Reducing the grammar

Definition

Let $\simeq \, = \, \preceq \cap \preceq^{-1}$. Let $(G/{\simeq}) = (N', \Sigma, \mathcal{S}, S', P')$

- $N' = (N/{\simeq})$
- $S' = [S]$
- for every $\sigma \in \Sigma_k$ and $A', A'_1, \ldots, A'_k \in N'$

$$\text{wt}(A' \to \sigma(A'_1, \ldots, A'_k)) = \sum_{A \in A'} \text{wt}(A \to \sigma(A_1, \ldots, A_k))$$

with $A_i \in A'_i$ for every $1 \leq i \leq k$

Theorem

G and $(G/{\simeq})$ are equivalent.

Reducing the grammar

Definition

Let $\simeq \; = \; \preceq \cap \preceq^{-1}$. Let $(G/\simeq) = (N', \Sigma, \mathcal{S}, S', P')$

- $N' = (N/\simeq)$
- $S' = [S]$
- for every $\sigma \in \Sigma_k$ and $A', A'_1, \ldots, A'_k \in N'$

$$\mathrm{wt}(A' \to \sigma(A'_1, \ldots, A'_k)) = \sum_{A \in A'} \mathrm{wt}(A \to \sigma(A_1, \ldots, A_k))$$

with $A_i \in A'_i$ for every $1 \leq i \leq k$

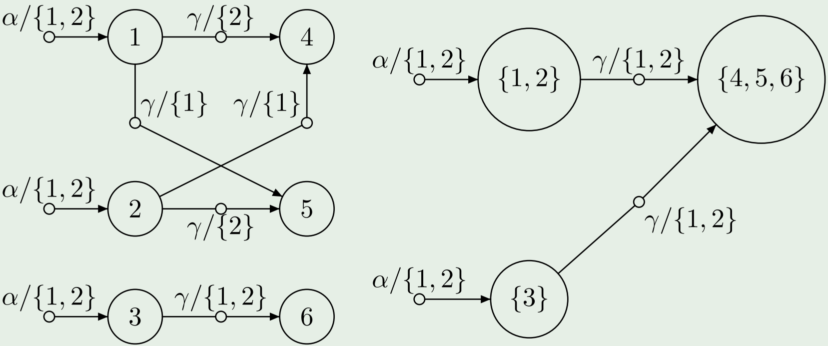
Theorem

G and $(G/\simeq)$ are equivalent.

Full example

Semiring $(\mathcal{P}(\{1,2\}), \cup, \cap, \emptyset, \{1,2\})$

Example (Old forward simulation)

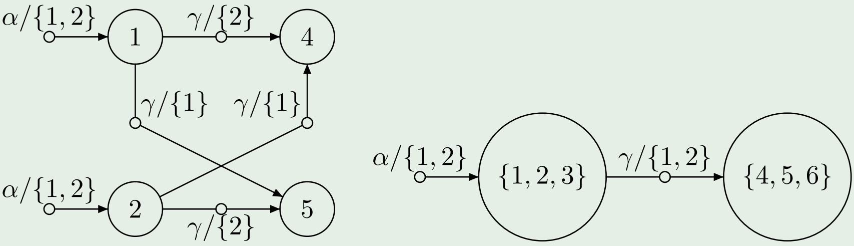


$1 \preceq 2 \preceq 1 \preceq 3$ and $4 \preceq 5 \preceq 6 \preceq 4$

Full example

Semiring $(\mathcal{P}(\{1,2\}), \cup, \cap, \emptyset, \{1,2\})$

Example (New forward simulation)



$1 \preceq 2 \preceq 3 \preceq 1$ and $4 \preceq 5 \preceq 6 \preceq 4$

Algorithm computing the greatest simulation

$R_0 \leftarrow \{(A, B) \in N \times N \mid \text{ if } A = S \text{ then } B = S\}$
$i \leftarrow 0$
repeat
  $j \leftarrow i$
  for all $\sigma \in \Sigma_k$, $1 \leq n \leq k$, and $A', A_1, \ldots, A_k \in N$ **do**
    $R_{i+1} \leftarrow \{(A, B) \in R_i \mid$

$$\sum_{B' \in \uparrow_{R_i} A'} \text{wt}(B' \to \sigma(\ldots, A, \ldots)) \sqsubseteq \sum_{B' \in \uparrow_{R_i} A'} \text{wt}(B' \to \sigma(\ldots, B, \ldots))\}$$

    $i \leftarrow i + 1$
  end for
until $R_i = R_j$

Additional (new) properties

Theorem

Every forward bisimulation [Högberg et al ’07] is a forward simulation.

Theorem

If $\preceq$ is the greatest forward simulation, then $(G/\simeq)$ with $\simeq = \preceq \cap \preceq^{-1}$ is forward-simulation minimal.

(i.e., it cannot be reduced further via forward simulation)

Additional (new) properties

Theorem

Every forward bisimulation [Högberg et al '07] is a forward simulation.

Theorem

If $\preceq$ *is the greatest forward simulation, then* $(G/\simeq)$ *with* $\simeq \,=\, \preceq \cap \preceq^{-1}$ *is forward-simulation minimal.*

(i.e., it cannot be reduced further via forward simulation)

State of the art

Property	*Simulation* New backward	New forward
Generalization		
unweighted simulation	✓	✓
weighted bisimulation	✓	✓
Computation		
Admits greatest simulation	✓	✓
Yields minimal device	✓	✓
Easy to compute	?	?

References

- ABDULLA, BOUAJJANI, HOLÍK, KAATI, VOJNAR
  Computing simulations over tree automata.
  Proc. TACAS, LNCS 4963, 2008
- BORCHARDT
  The theory of recognizable tree series.
  PhD thesis, TU Dresden, 2005
- HÖGBERG, MALETTI, MAY
  Backward and forward bisimulation minimization of tree automata.
  Theoret. Comput. Sci., to appear.
- MALETTI
  A backward and a forward simulation for weighted tree automata.
  Proc. CAI, LNCS 5725, 2009

Thank you for your attention!

References

- ABDULLA, BOUAJJANI, HOLÍK, KAATI, VOJNAR
  Computing simulations over tree automata.
  Proc. TACAS, LNCS 4963, 2008
- BORCHARDT
  The theory of recognizable tree series.
  PhD thesis, TU Dresden, 2005
- HÖGBERG, MALETTI, MAY
  Backward and forward bisimulation minimization of tree automata.
  Theoret. Comput. Sci., to appear.
- MALETTI
  A backward and a forward simulation for weighted tree automata.
  Proc. CAI, LNCS 5725, 2009

Thank you for your attention!